Evolutionary Approach for Selection of Optimal EEG Electrode Positions and Features for Classification of Cognitive Tasks

Rimita Lahiri, Pratyusha Rakshit, Amit Konar
Electronics & Telecommunication Engineering Department
Jadavpur University, Kolkata, India
rims.92@gmail.com, pratyushar1@gmail.com, konaramit@yahoo.co.in

Atulya K. Nagar
Department of Mathematics and Computer Science
Liverpool Hope University
Liverpool, UK
nagara@hope.ac.uk

***Abstract*—This paper proposes a novel evolutionary approach to the optimal selection of electrodes as well as relevant EEG features for effective classification of cognitive tasks. The problem has been formulated in the framework of a single objective optimization problem with an aim to simultaneously satisfying three criteria. The first criterion deals with maximization of the correlation between the features of EEG sources before and after the selection of optimal electrodes. The second criterion is concerned with minimization of the mutual information between the features of the selected EEG electrodes. The last criterion aims at maximization of the ratio of the difference between the selected features of the EEG sources between and within any two cognitive tasks. A self-adaptive variant of FA (referred to as SAFA) is proposed to solve the above optimization problem by proficiently balancing the trade-off between the computational accuracy and the run-time complexity. Experiments undertaken over wide variety of cognitive tasks reveal that the proposed algorithm outperforms the other standard algorithms (applied to the same problem) in terms of accuracy and computational overhead.**

Keywords—EEG electrodes; source signal; sink signal; EEG feature; firefly algorithm.

I. INTRODUCTION

Brain computer interfacing (BCI) [1] is a multi-dimensional field of research, concerned with cognition, neurophysiology, psychology, sensors, machine learning, signal detection and processing, to name a few. Now a day, BCI stands alone as the only modality of control and communication for patients suffering from diseases like amyotrophic lateral sclerosis, paralysis, cerebral palsy, and amputees. Its contributions in medical fields range from prevention to neuronal rehabilitation for serious injuries. BCI addresses analyzing, conceptualization, monitoring, measuring, and evaluating the complex neuro-physiological behaviors detected and extracted from a set of electrodes over the scalp or from those implanted inside the brain.

These BCI interfaces bypass the natural pathways of neuro-muscular control and thus aim at serving an alternative means of communication/control in case of failure in neural/motor functioning. Several interfacing methodologies including invasive implants, semi invasive implants like electrocorticography (ECoG) and non invasive modalities like electroencephalogram (EEG), magnetoencephalogram (MEG) and functional magnetic resonance imaging (fMRI) have emerged in order to implement BCI successfully. EEG is the preferred technology for measuring brain activities for most BCI researchers because of its non-invasiveness, portability, easy availability, and high temporal resolution.

The basic BCI module consists of three steps, including i) pre-processing of the EEG signals dealing with artifact removal, identification of relevant electrodes and frequency bands of EEG signals, ii) feature extraction, and iii) classification, concerned with identification of different mental states. The classified results lead to the generation of the control signals required to drive an assistive device. The classification accuracy relies on the extent of detour the redundant information. This paper addresses two crucial factors for effective classification of cognitive tasks using EEG based BCI systems, including

1. Optimal selection of electrodes [5] to facilitate faster processing of EEG signals for different cognitive tasks, and
2. Optimal selection of relevant EEG features to enhance the performance of a classifier.

The optimal selection of electrodes [5] is essentially influenced by the estimation of cortical sources. The EEG devices acquire raw cortical current signals, generated from different independent sources, through neuronal firing in outer cortex of the brain. These signals are then transformed to respective voltage signals by passing through different resistive devices. Finally, the voltage signals are recorded by placing electrodes at specified scalp regions. Due to volume conduction, the signal acquired at the scalp electrodes is found to contain components of different sources. Moreover, for a particular cognitive task every electrode placed on the scalp cannot provide relevant information, in fact at times electrodes generate redundant information. Hence, optimal electrode selection is very important for signal analysis and relevant decision making in the following steps. Otherwise overlapping information will not only degrade performance metrics, but it involves processing of same signal components more than once, which negatively affects the time complexity as well.

Only optimal electrode selection is not sufficient for a successful EEG based BCI implementation. One of the significant concerns in BCI research is to deal with the high dimensionality of the features. Often it is observed that due to the presence of a large number of redundant features in the feature set, the accuracy of the classifier is greatly reduced. Researchers are now taking keen interest to select fewer discriminate features from the high dimensional EEG feature vectors for different cognitive tasks without sacrificing the classification accuracy. The paper proposes a novel evolutionary approach to automatic selection of optimal set of EEG electrodes and EEG features (from the high dimensional feature space). The principle of evolutionary electrode and feature selection is outlined next.

In this paper, the possible selection of EEG electrodes is realized by optimizing the scoring function, which deals with

1. maximizing the association between the estimated source signals corresponding to the original set and the reduced set of electrodes, and

2. minimizing the mutual information between two selected electrodes.

The first criterion reduces the loss in information of the cortical sources corresponding to a cognitive task after reducing the number of electrodes. The second criterion aims at identifying the relevant electrodes conveying unique information of specific cognitive tasks, thus discarding the redundant information.

The design philosophy adopted for the optimal selection of EEG features is to identify the set of features that are capable

1) to uniquely represent a specific class of cognitive tasks and

ii) to effectively differentiate between any two classes.

It is realized by jointly serving the following two criteria. First, the selected j-th features of the data points in a given class should be close to each other. Contrarily, the difference between the means of the selected j-th feature of any two classes should be as high as possible.

A self-adaptive variant of the traditional firefly algorithm (FA) [6] is proposed here to select an optimal set of appropriate EEG electrodes and features by jointly optimizing the above-mentioned objectives. FA is selected here partly heuristically and partly due to its established performance with respect to computational accuracy and run-time complexity [6]. The self-adaptive variant of FA assists the potential candidate solutions (of the optimization problem) to confine their search in their local neighborhood in the parameter space. On the other hand, the inferior members are equipped with global exploration capability.

The present work aims at improving the work proposed in [3] in three important aspects. First, the work proposed in [3] is primarily concerned with the optimal electrode selection for a specific cognitive task. However, in this paper, we aim at selecting the optimal electrodes for effective classification of different cognitive tasks. This is a more realistic scenario of practical BCI applications. Second, Pearson correlation coefficient is utilized in [3] to capture the degree of relationship between the information conveyed by the original set of electrodes and the reduced number of selected electrodes. This correlation measure is found to be only sensitive to the linear relationship between two variables, incompetent to model non-linear or monotonous relationships. In other words, a zero value of Pearson coefficient does not always imply independence between two variables. Thus, it may fail to proficiently capture the dependence between two EEG sources. This is here circumvented by using an alternative measure of correlation, referred to as the distance correlation coefficient [7]. Third, the present work also attempts to select the optimal selection of EEG features along with the EEG electrodes. This improves the classification accuracy.

The rest of the paper is divided into five sections. Section II provides a detailed description of the proposed framework. Section III recapitulates the traditional FA and presents the proposed self-adaptive FA (SAFA). The experimental results are reported in section IV. Section V concludes the paper.

II. PROPOSED METHODOLOGY

This section provides a detailed overview of the proposed framework (Fig. 1). The proposed method involves a training data set **T**, which is pictorially illustrated in Fig. 2, for a specific cognitive class K_c. The training dataset consists of L data points, each having information of N electrode (sink) signals. Each sink signal is represented as a F-dimensional feature vector. The paper aims at the optimal selection of M ($\leq N$) electrodes and D ($\leq F$) features for effective classification of C cognitive tasks. It is to worth mentioning that K_c=[1, C]. Here $S_{i,k}^{t,c}$ and $S_{i,k}^{t,c}$ denote the k-th feature of the i-th electrode and the corresponding source for the t-th data point in the class K_c.

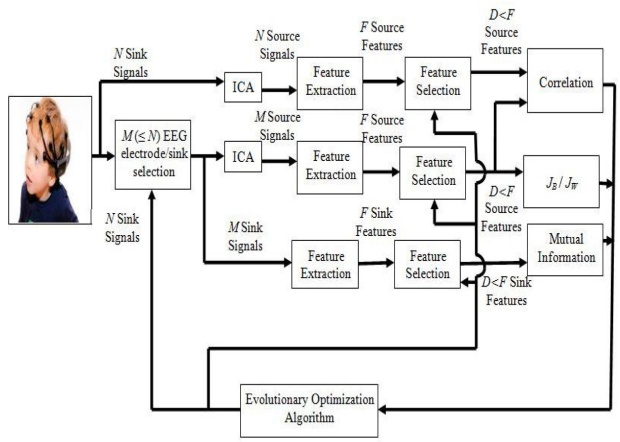


Fig. 1 The overview of the proposed scheme

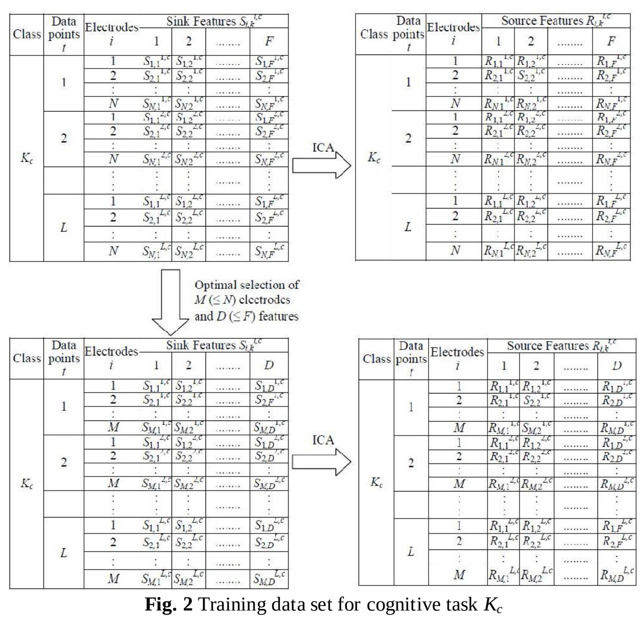


Fig. 2 Training data set for cognitive task K_c

A. Independent Component Analysis as a Source Localization Tool

There are many sources inside human brain, which produce current signals due to neuronal firing during any cognitive task. The voltage signals recorded by the EEG electrodes (also referred to as sink signals) placed on a human scalp are essentially mixtures of these source signals. In brain imaging, the problem of estimating the locations and distributions of the cortical sources, based on the voltage readings of the EEG electrodes, remains as an 'ill-posed' blind signal separation problem. Independent component analysis (ICA) provides a solution to this problem [3] by transforming a multivariate signal (for instance, EEG signal) into a linear combination of independent non-Gaussian subcomponents (for example, non-Gaussian source signals).

Let, $\mathbf{P} = [\vec{P}_1, \vec{P}_2, ..., \vec{P}_N]^T$ and $Q = [\vec{Q}_1, \vec{Q}_2, ..., \vec{Q}_N]^T$ be the N source and sink signals in time domain. ICA deals with representing each $\vec{Q}_i$ as a linear combination of $\vec{P}_j$ for j=[1, N], given by

$$\vec{Q}_i = w_{i,1}\vec{P}_1 + w_{i,2}\vec{P}_2 + ... + w_{i,N}\vec{P}_N = \sum_{j=1}^{N} w_{i,j}\vec{P}_j \quad \text{for } i=[1, N] \tag{1}$$

or

$$\mathbf{Q} = \mathbf{WP} \tag{2}$$

with $w_{i,j}$ representing the degree of dependence of $\vec{Q}_i$ on $\vec{P}_j$ based on their distance and **W** denoting the mixing matrix with elements for i, j=[1, N]. ICA aims at determining the source signals by identifying the optimal mixing matrix, obtained by the minimization of the mutual information or the maximization of the non-Gaussianity of the estimated source signals. The members of former family use measures like K-L divergence and maximum entropy while the latter family utilizes kurtosis, negentropy and so on. In the present paper, we employed a variant of ICA based on the Infomax algorithm [3] that uses the maximum entropy measure.

Once the mixing matrix **W** is estimated, the source signals can be derived by multiplying the observed sink signals with the inverse of the mixing matrix, also known as demixing matrix $\mathbf{V}=\mathbf{W}^{-1}$, by

$$\mathbf{P} = \mathbf{W}^{-1}\mathbf{Q} = \mathbf{VQ} \tag{3}$$

B. Feature Extraction

EEG signals of the sinks for two different cognitive tasks, as well as the sources, have been represented using five well-known features. The feature set consists of common spatial feature (CSP) [2], two time domain features, including adaptive autoregressive (AAR) Parameters [3], and Hjorth parameters [3], a frequency domain feature, including power spectral density (PSD) [3] and a set of time-frequency correlated features, including discrete wavelet coefficients [3].

C. Optimum Features and Electrode Selection using Evolutionary Approach

The performance of any real world optimization algorithm, as in the preset context, greatly relies on the judicial formulation of the objective function. In the present scenario, the optimal selection of M ($\leq N$) EEG electrodes and D ($\leq F$) salient EEG features for classification of two cognitive tasks is realized based on three significant performance characteristics. The first two characteristics deal with the optimal selection of electrodes, while the third one is concerned with the selection of optimal feature.

1. For a specific cognitive class K_c and t-th data point, the correlation between the D features of the source signals estimated from the optimally selected M electrodes and that of the original N electrodes should be maximized.

One of the major objectives in the present context is to preserve the relevant information of the recorded EEG signals for accurate classification of cognitive tasks, even after selecting the optimal set of M electrodes out of N. A higher degree of association between the sets of source features corresponding to the original N and the selected M ($\leq N$) sink signals indicate an efficient representation of the original number (N) of sink signals with the reduced number (M) of sinks signals. The degree of relationship here is captured by *distance correlation* [7] between the feature space of M and N source signals. The distance correlation is an index in [0, 1] representing the statistical dependence between two sets of source features.

Distance correlation between any two sets of D-dimensional source features, $\vec{R}_i^{t,c}$ and $\vec{R}_j^{t,c}$, (for t-th data point of cognitive class K_c) is determined by

$$\phi_{i,j}^{t,c} = \frac{dC_{i,j}}{\sqrt{dV_i \times dV_j}} \tag{4}$$

where
$$dC_{i,j} = \frac{1}{D}\sqrt{\sum_{k,l=1}^{D} A_{k,l}B_{k,l}} \tag{5}$$

and
$$\left.\begin{aligned} dV_i &= \frac{1}{D^2}\sum_{k,l=1}^{D} A_{k,l}^2 \\ dV_j &= \frac{1}{D^2}\sum_{k,l=1}^{D} B_{k,l}^2 \end{aligned}\right\} \tag{6}$$

with the components of the matrix **A** being determined by

and
$$\left.\begin{aligned} a_{k,l} &= \left\| R_{i,k}^{t,c} - R_{i,l}^{t,c} \right\|, \quad \text{for } k,l = [1,F] \\ A_{k,l} &= a_{k,l} - \bar{a}_k - \bar{a}_l + \bar{a} \quad \text{for } k,l = [1,F] \\ \bar{a}_k &= \frac{1}{F}\sum_{l=1}^{F} a_{k,l}, \quad \bar{a}_l = \frac{1}{F}\sum_{k=1}^{F} a_{k,l} \\ \bar{a} &= \frac{1}{F^2}\sum_{k,l=1}^{F} a_{k,l}. \end{aligned}\right\} \tag{7}$$

The components of matrix **B** are determined similarly as in (7), however, using the features of $\vec{R}_j^{t,c}$. Here ||.|| denotes the Euclidean norm. A value of $\phi_{i,j}$=0 is an implication of independence of $\vec{R}_i^{t,c}$ and $\vec{R}_j^{t,c}$. Distance coefficient is here adopted as a measure of correlation between source features, instead of Pearson correlation coefficient, to overcome its incapability to characterize non-linear or monotone dependency [7]. An accurate selection of optimal electrodes will increase the distance correlation, indicating that the source features corresponding to the original and reduced set of electrodes co-vary with each other.

2. For a specific cognitive class K_c and t-th data point, the mutual information between the D features of two selected electrodes should be minimized.

From the point of view of the present problem, in order to maximize the unique information content of two maximally distinct selected sinks, the redundant information is to be reduced. There are possibilities that for a particular feature, any two electrodes residing in close proximity are providing similar information. Therefore, inclusion of any one between the two electrodes (instead of considering both) will serve the purpose with reduced computational complexity. The similarity in the information content of two sink feature sets is modeled here by their mutual information (MI).

Based on information theory, MI is used to identify the amount of uncertainty about one sink feature vector $\vec{S}_i^{t,c}$ given knowledge of the other sink feature vector $\vec{S}_j^{t,c}$ (for cognitive class K_c and t-th data point). It equals zero if they are independent. Thus, more the mutual information between them, less the uncertainty in $\vec{S}_i^{t,c}$ given the knowledge of $\vec{S}_j^{t,c}$ or vice versa and hence, only one of them should be selected. Hence $MI_{i,j}^{t,c}$ is utilized here as a dependence measure of $\vec{S}_i^{t,c}$ on $\vec{S}_j^{t,c}$ and vice versa, given by

$$MI_{i,j}^{t,c} = H_i - H_{i|j} \tag{8}$$

$$H_i = -\sum_{k=1}^{D} p_i(k) \times \log_2 p_i(k) \tag{9}$$

$$H_{i|j} = -\sum_{k=1}^{D} p_{i,j}(k) \times \log_2 p_{i|j}(k) \tag{10}$$

where, the average information or entropy H_i is the uncertainty in $\vec{S}_i^{t,c}$ before observing $\vec{S}_j^{t,c}$ and the conditional entropy $H_{i|j}$ represents the uncertainty in $\vec{S}_i^{t,c}$ after observing $\vec{S}_j^{t,c}$. Here $p_i(k)$ is the probability of each feature in $\vec{S}_i^{t,c}$, $p_{i,j}(k)$ is the joint probability of $\vec{S}_i^{t,c}$ and $\vec{S}_j^{t,c}$ and $p_{i|j}(k)$ is the transition probability from $\vec{S}_i^{t,c}$ to $\vec{S}_j^{t,c}$.

3. For a specific cognitive task K_c, the similarity between the M source signals represented by the selected D features should be high. Contrarily, for better separability between two cognitive tasks, the difference between their respective source feature vectors should be maximized.

This objective is concerned with the optimal selection of D features out of F features of the source signals corresponding to the selected M ($\leq N$) sink signals. This aims at discarding the redundant information of the selected source signals for classifying two specific cognitive tasks. Let L be the number of data points (source EEG feature vectors) of any cognitive task.

This is accomplished here in two phases. First, it tries to minimize the difference between the selected D ($\leq F$) features of the source signals (corresponding to the selected $M \leq N$ sink signals), for all data points within a specific class K_c. The design philosophy is that if the k-th feature is selected as a unique representative feature of the class K_c, then it should provide a high degree of similarity between the k-th feature of the i-th source for any two data points within the same class K_c. This is done for all classes c=[1, C]. This is realized here by minimizing

$$J_W = \sum_{c=1}^{C} \sum_{t=1}^{L-1} \sum_{u=t+1}^{L} \sum_{i=1}^{M} \sum_{k=1}^{D} \left(R_{i,k}^{t,c} - R_{i,k}^{u,c} \right)^2. \tag{11}$$

The second part aims at maximizing the ratio of the mean to the standard deviation between the selected D ($\leq F$) features of the source signals (corresponding to the selected $M \leq N$ sink signals), between any two classes. This ensures that the selected feature is capable enough to discriminate between any two classes. This is here realized by maximizing

$$J_B = \sum_{c=1}^{C} \sum_{\substack{d=1,\\ d \neq c}}^{C} \sum_{k=1}^{D} \sum_{i=1}^{M} \left(\bar{R}_{i,k}^{c} \big/ \sigma_{i,k}^{c} - \bar{R}_{i,k}^{d} \big/ \sigma_{i,k}^{d} \right) \tag{12}$$

where
$$\bar{R}_{i,k}^{c} = \frac{1}{L} \sum_{t=1}^{L} R_{i,k}^{t,c} \tag{13}$$

$$\sigma_{i,k}^{c} = \sqrt{\frac{(R_{i,k}^{t,c} - \bar{R}_{i,k}^{c})^2}{L}} \tag{14}$$

represent the mean and the standard deviation of the k-th feature of the i-th electrode over L data points in a specific class K_c for k=[1, D], i=[1, M] and K_c=[1, C].

With these three considerations, a composite objective function is formulated in (15), maximization of which yields the optimal sets of electrodes and features for classification of C cognitive tasks.

$$f = \sum_{c=1}^{C} \sum_{t=1}^{L} \sum_{i=1}^{N} \sum_{\substack{j=1,\\ j \neq i}}^{M} \phi_{i,j}^{t,c} - \lambda_1 \sum_{c=1}^{C} \sum_{t=1}^{L} \sum_{i=1}^{M} \sum_{\substack{j=1,\\ j \neq i}}^{M} MI_{i,j}^{t,c} + \lambda_2 \frac{J_B}{J_W} \tag{15}$$

λ_1 and λ_2 are constants which are set in a manner so as to have all terms in the right hand side of (15) in the same order of magnitude. In our experiment, λ_1 and λ_2 are respectively set as 10 and 20.

Fig. 3 pictorially represents a four dimensional candidate solution $\vec{Y}$ for the present optimization problem with N=5 electrodes and F=4 features. Y_1 and Y_3 denote the values of M ($\leq N$) and D ($\leq F$). Y_2 and Y_4 are decimal values within [1, 2^N-1] and [1, 2^F-1] respectively. The binary decoding of Y_2 and Y_4 are used to identify M electrodes and D features. For example, let $\vec{Z}_2$ (or $\vec{Z}_4$) be N (or F) dimensional binary string, obtained by decoding Y_2 (or Y_4). Intuitively, $Z_{2,j}$= 1 ($Z_{4,j}$= 1) indicates that the j-th electrode (or feature) selected or $j \in$[1, N] (or $j \in$[1, F]). It is noteworthy that the number of ones in $\vec{Z}_2$ (or $\vec{Z}_4$) should be equal to M (or D).

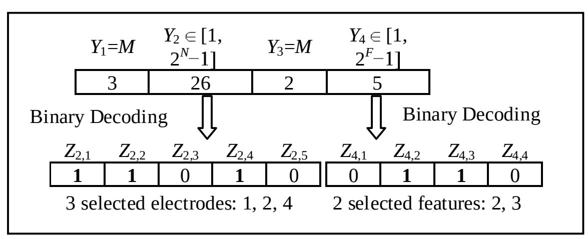


Fig. 3 Illustration of encoding a candidate solution for N=5 and F=4

D. Classification

Classification is considered as an important step in EEG based research problems. A classifier primarily segregates unknown data samples into considered class labels after being trained with similar features. In context of BCI research, a classifier aims to distinguish between different brain activities after being optimally trained by following any 'learning' algorithm. While classifying EEG signals obtained from the electrodes placed on the scalp, every researcher strives to achieve high accuracy. There are numerous kinds of classifiers available which lets a user to choose the most suitable classifier according to the requirement of the problem. In the present paper, we have used non-linear *support vector machine* (SVM) classifier with suitable kernel function for each of the tasks conducted. Three kernel functions, including

1. Gaussian radial basis function:
   $k(X_1, X_2) = \exp(-\lambda \|X_1 - X_2\|^2)$ for $\lambda > 0$
2. Homogeneous polynomial function:
   $k(X_1, X_2) = (X_1.X_2)l$ where l denotes the number of polynomials
3. Hyperbolic tangent function:
   $k(X_1, X_2) = \tanh(kX_1.X_2 + c) \in k < 0$ and $c > 0$.

To determine the efficacy of the proposed methodology, the SVM classifier is trained with the selected features of the selected electrodes (and corresponding sources). The testing data set is also prepared with the selected feature-based information contents of the selected electrodes and the sources.

III. SELF-ADAPTIVE FIREFLY ALGORITHM

The optimization problem of optimal selection of electrode positions and EEG features for classification of cognitive tasks here has solved using a self-adaptive variant of the traditional firefly algorithm (FA). In this section, first an overview of the traditional FA is provided. Next, we propose the self-adaptive variant of FA, referred to as SAFA, which adaptively tunes its control parameters to balance the trade-off between the computational accuracy and the run-time complexity effectively.

A. Firefly Algorithm

In firefly algorithm (FA) [6], a possible solution of an optimization problem is encoded by the position of a firefly in the parameter space and its light intensity signifies the fitness of the respective solution. An overview of FA is given next.

1. Initialization: Initially, a population $\mathbf{P}(t)$ of NP, D-dimensional firefly positions, $\vec{Y}_i(t) = \{y_{i,1}(t), y_{i,2}(t), \ldots, y_{i,D}(t)\}$ for i= [1, NP] is uniformly randomized in the search range $[\vec{Y}^{\min}, \vec{Y}^{\max}]$ where $\vec{Y}^{\min} = \{y_1^{\min}, y_2^{\min}, \ldots, y_D^{\min}\}$ and $\vec{Y}^{\max} = \{y_1^{\max}, y_2^{\max}, \ldots, y_D^{\max}\}$ at the current generation t = 0, given by

$$y_{i,d}(0) = y_d^{\min} + \text{rand}(0,1) \times (y_d^{\max} - y_d^{\min}) \tag{16}$$

for j=[1, D] and i=[1, NP] where rand(0, 1) is a uniformly distributed random number in [0, 1]. The objective function value $f(\vec{Y}_i(0))$ of $\vec{Y}_i(0)$ is evaluated for i= [1, NP].

2. Attraction to Brighter Fireflies: Now the firefly $\vec{Y}_i(t)$ is attracted towards the positions of the brighter fireflies $\vec{Y}_j(t)$ for i, j= [1, NP] but $i \neq j$ such that $f(\vec{Y}_j(t)) > f(\vec{Y}_i(t))$ (for maximization problem). Apparently, the attractiveness $\beta_{i,j}$ of $\vec{Y}_i(t)$ towards $\vec{Y}_j(t)$ decreases exponentially with the distance between them, denoted by $d_{i,j}$ as given in (17).

$$\beta_{i,j} = \beta_o \exp(-\gamma \times d_{i,j}^m), \quad m \geq 1 \tag{17}$$

where β_0 represents the maximum attractiveness felt by $\vec{Y}_i(t)$ at its own position (i.e., at $d_{i,j} = d_{i,i} = 0$) and γ denotes the light absorption coefficient, which controls the rate of change of $\beta_{i,j}$ with $d_{i,j}$. Intuitively, γ governs the convergence speed of FA [6]. A setting of γ=0 is concerned with a constant attractiveness of β_0 for all fireflies, while γ approaching infinity implies complete random search [6]. The possible range of γ is found to be [0.01, 10] in the existing literature. In (17), m is a pre-defined positive non-linear modulation index. The distance between $\vec{Y}_i(t)$ and $\vec{Y}_j(t)$ is computed using the Euclidean norm as follows.

$$d_{i,j} = \| \vec{Y}_i(t) - \vec{Y}_j(t) \| \tag{18}$$

This step is repeated for i, j= [1, NP].

3. Movement of Fireflies: The firefly at position $\vec{Y}_i(t)$ flies towards a more attractive location $\vec{Y}_j(t)$ (in the parameter space) of a brighter firefly (i.e., $f(\vec{Y}_j(t)) > f(\vec{Y}_i(t))$) for j= [1, NP] but $i \neq j$ following

$$\begin{aligned} &\vec{Y}_i^{next}(t) = \vec{Y}_i^{cur}(t) + \beta_{i,j} \times (\vec{Y}_j(t) - \vec{Y}_i(t)) + \alpha \times (\vec{r} - 0.5) \\ &\vec{Y}_i^{cur}(t) \leftarrow \vec{Y}_i^{next}(t) \end{aligned} \tag{19}$$

where $\vec{Y}_i^{cur}(t)$ is initialized with $\vec{Y}_i(t)$ before its movement and $\vec{r}$ denotes a D-dimensional random position vector with its d-th component uniformly distributed in [0, 1] for d=[1, D]. The movement of the i-th firefly, governed by (19), is carried on for j=[1, NP], but $i \neq j$ such that $f(\vec{Y}_j(t)) > f(\vec{Y}_i(t))$. This step is repeated for i= [1, NP]. The first term in (19) represents the firefly's position after its last movement. The second term in (19) denotes the positional change of $\vec{Y}_i(t)$ due to the attraction towards $\vec{Y}_j(t)$. Apparently, this term has o contribution towards controlling the movement of the brightest firefly and hence, it may be stuck at the local optima in the parameter space. This problem is overcome by inducing a random movement of the fireflies with a step-size of $\alpha \in$ (0, 1). The upgraded position of the i-th firefly, after being controlled by the brighter ones, is represented by $\vec{Y}_i(t+1)$ for i= [1, NP].

4. Convergence: After each evolution, the steps 2 and 3 are repeated until one of the terminating conditions is satisfied. The conditions include restricting the number of maximum generations, preserving error limits, or the both, whichever occurs earlier.

B. Self-Adaptive Firefly Algorithm (SAFA)

The population members of the traditional FA are equipped with the exploitation capability capable to escape the local optima due to their random movements with step-size α in (19). Intuitively, the step-size (α) profile governs the convergence of fireflies towards global optimum. However, in traditional FA, a constant value of α is used for the random movement, irrespective of the position of the fireflies in the parameter space. A setting of large value of α may result in the deviation of a quality firefly from the global optimum, while a small value of α may take a relatively long time to effectively orient a poor firefly towards the global optimum. The exploitation capability of the traditional FA, being a decisive factor of its performance, here has been farther improved by self-adapting the step-size parameter α within a range $[\alpha^{\min}, 1)$ based on the relative position of a firefly with respect to the current best firefly position. This is realized by setting

$$\alpha_{i,d} = \alpha^{\min} + (1-\alpha^{\min}) \times \mathrm{rand}(0,1) \times \frac{|\, y_d^{best}(t) - y_{i,d}(t)\,|}{y_d^{\max} - y_d^{\min}} \quad (20)$$

for d=[1, D]. The step-size is now treated as a D-dimensional vector as symbolized by $\vec{\alpha}_i = \{\alpha_{i,1}(t), \alpha_{i,2}(t), \ldots, \alpha_{i,D}(t)\}$ with its d-th component $\alpha_{i,d} \in [\alpha^{\min}, 1)$ for d=[1, D]. Here $\vec{Y}^{best}(t) = \{y_1^{best}(t), y_2^{best}(t), \ldots, y_D^{best}(t)\}$ the best position of the firefly in the t-th generation and |.| represents the absolute value. The dynamic in (20) guarantees that a firefly at $\vec{Y}_i(t)$, close to $\vec{Y}^{best}(t)$, should exploit the local neighborhood with a small step-size $\alpha_{i,d} \approx \alpha^{\min}$ to prevent the exclusion of the global optimum. Contrarily, an inferior firefly, far away from $\vec{Y}^{best}(t)$, should take part in the global search (with step size $\alpha_{i,d}$ approaching unity for d=[1, D]) to explore the potential zones in the parameter space.

Moreover, a simple strategy is proposed to update the values of γ in each generation, based on the knowledge of its potential values that were able to generate better firefly positions in the last generation. At every generation, the light absorption coefficient $\gamma_i(t)$ of each individual firefly $\vec{Y}_i(t)$ is independently generated as

$$\gamma_i(t) = Rayleigh(\sqrt{2/\pi}\,\bar{\gamma}(t)) \quad (21)$$

where $Rayleigh(\sqrt{2/\pi}\,\bar{\gamma}(t))$ is a random number sampled from a Rayleigh distribution with mean $\bar{\gamma}(t)$ (scale parameter $\sqrt{2/\pi}\,\bar{\gamma}(t)$). The value of $\gamma_i(t)$ is reproduced if it is beyond its allowable range [0.01, 10].

Let $\gamma_s(t)$ be the set of the successful light absorption coefficients of all fireflies of the current generation t producing better positions for the next generation t+1.

$$\gamma_s(t) = \{\gamma_i(t) \text{ for } i = [1, NP] : f(\vec{Y}_i(t+1)) > f(\vec{Y}_i(t))\} \quad (22)$$

Here, $\bar{\gamma}(0)$ is initialized to be 0.5. After every generation, it is updated as

$$\bar{\gamma}(t+1) = w \times \bar{\gamma}(t) + (1-w) \times \mu_p(\gamma_s(t)) \quad (23)$$

where $\mu_{pow}(.)$ denotes the power mean [8], given by

$$\mu_p(\gamma_s(t)) = \sum_{\gamma \in \gamma_s(t)} [\gamma^n / |\gamma_s(t)|]^{1/n} \quad (24)$$

with $|S|$ denotes the cardinality of set S. The weight factor w in (23) is randomly selected from [0.7, 1]. The weighted sum of $\bar{\gamma}(t)$ and $\mu_p(\gamma_s(t))$ helps in the effective tuning of $\bar{\gamma}(t+1)$ based on the successful values of the light absorption coefficients in the past and present generations respectively. The positive constant n in (24) is taken as 1.5 after wide variety of experiments to avoid premature convergence at local optima. The design philosophy adopted here relies on the principle of selecting large diversified values of γ from the Rayleigh distribution (having longer tails than the normal distribution) when the population is far away from the global optimum.

Procedure SAFA_Induced_Electrode_Feature_Selection

Input: L data points, each comprising the EEG signals of N electrodes for each cognitive task class K_c for c=[1, C], F EEG features, non-linear modulation index m.

Output: Optimally selected M ($\le N$) electrodes and D ($\le F$) features.

Begin

1. Initialize a population $\mathbf{P}(t)$ of NP, 4-dimensional firefly position vectors $\vec{Y}_i(t)$ at generation t=0 using (16) and Fig. 3 for i= [1, NP].
2. Set $\bar{\gamma}(0) \leftarrow 0.5$.
3. Decode $\vec{Y}_i(0)$ using Fig. 3 and evaluate $f(\vec{Y}_i(0))$ using (15) for i= [1, NP].
4. Set $\vec{Y}^{best}(0) \leftarrow \arg\left(\max_{i=1}^{NP} f(\vec{Y}_i(0))\right)$.
5. **While** termination condition is not reached **do**

   Begin

   I. Set $\gamma_s(t) \leftarrow NULL$.

   II. **For** i=1 to NP **do**

   Begin

   (i) Select $\gamma_i(t)$ using (21).

   (ii) **For** j=1 to NP, $j \ne i$ **do**

   Begin

   If $f(\vec{Y}_j(t)) > f(\vec{Y}_i(t))$ **Then do**

   Determine $\vec{Y}_i^{next}(t)$ using (19) and (20).

   End If.

   End For.

   (iii) Set $\vec{Y}_i(t+1) \leftarrow \vec{Y}_i^{next}(t)$.

   (iv) Decode $\vec{Y}_i(t+1)$ using Fig. 3 and evaluate $f(\vec{Y}_i(t+1))$ using (15) for i= [1, NP].

   (v) **If** $f(\vec{Y}_i(t+1)) > f(\vec{Y}_i(t))$ **Then**

   Set $\gamma_s(t) \leftarrow \gamma_s(t) \cup \gamma_i(t)$.

   End If.

   End For.

   III. Set $\vec{Y}^{best}(t) \leftarrow \arg\left(\max_{i=1}^{NP} f(\vec{Y}_i(t))\right)$.

   IV. Set $t \leftarrow t$+1.

   V. Determine $\bar{\gamma}(t+1)$ from (23).

   End While.
6. Decode $\vec{Y}^{best}(t)$ using Fig. 3.
7. Return optimally selected M ($\le N$) electrodes and D ($\le F$) EEG features.

End.

TABLE I: EXPERIMENTS CONDUCTED FOR EXPERIMENTS

Index	Description
Experiment 1	Left hand and right hand motor execution
Experiment 2	Left hand and right hand motor imagery
Experiment 3	Left leg and right leg motor execution
Experiment 4	Left leg and right leg motor imagery
Experiment 5	Tongue and finger motor execution
Experiment 6	Happiness and sadness emotion recognition
Experiment 7	Happiness and anger emotion recognition
Experiment 8	Happiness and fear emotion recognition
Experiment 9	Happiness and disgust emotion recognition
Experiment 10	Disgust and fear emotion recognition
Experiment 11	Disgust and anger emotion recognition
Experiment 12	Disgust and sadness emotion recognition
Experiment 13	Anger and fear emotion recognition
Experiment 14	Anger and sadness emotion recognition
Experiment 15	Sadness and fear emotion recognition

IV. EXPERIMENTAL RESULTS AND PERFORMANCE ANALYSIS

This section describes each step of the proposed framework with the obtained readings arranged in a tabulated manner. Apart from that, the performance of the proposed system has been analyzed with respect to certain standard methods in terms of different performance indices.

A. EEG Signal Acquisition

Fifteen different binary classification experiments are carried out for this particular framework, in order to validate the obtained results for any kind of cognitive task without loss of generality. In each case, ten healthy subjects aged between 22 and 30 years have participated (five male and five female) in the experiments. The experiments are undertaken in multiple sessions, each of 2 minutes duration and every subject is asked to perform each session 10 times. It is thus evident from Fig. 2 that the number of data points L in each cognitive task equals to (5+5)×10=100. All the signals are recorded using a stand-alone EEG machine manufactured by Nihon Kohden (200 Hz sampling frequency) comprising N=19 electrodes, placed using the standard international 10-20 electrode placement method, as shown in Fig. 4. A_1 and A_2 are considered as the reference electrodes.

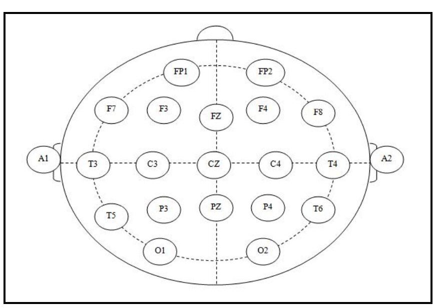


Fig. 4 Electrode positions according to 10-20 electrode placement system

Table-I presents an overview of the cognitive tasks undertaken during different experiments for the concerned problem. From Table-I, it can be seen that the experiments are not limited to only motor execution or motor intention based tasks, but the list also includes several experiments of segregation of five basic emotions, including happiness, sadness, anger, fear and disgust. To conduct experiments related to emotion recognition, certain video clips are shown to the users with sufficient amount of relaxation interval in between two clips corresponding to two different emotions. The clips of small duration are chosen such that it takes minimal time for the user to understand the clip and to experience such emotions freely.

While performing the experiments the subjects are asked to sit on a chair with arms at rest position and eyes stable. After a certain interval, a cross appears on the screen along with a beep sound; thereafter the subjects are instructed to perform the required tasks according to commands appearing on screen. The structure of a stimulus used for left and right hand motor execution task has been shown in Fig. 5 as an example.

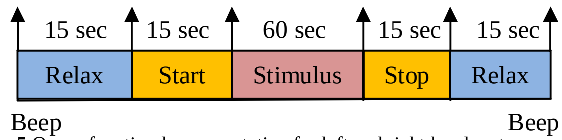


Fig. 5 Queue for stimulus presentation for left and right hand motor execution

B. Preprocessing

The acquired EEG signals are passed through certain preprocessing steps in order to remove the artifacts generated due to eye blinking or spurious pick-ups from the power supply. The responses to the experiments undertaken are better captured by the rhythmic brain activity in the frequency band of 4-40Hz. Hence, to filter the signals within the specified frequency band, a 6-th order elliptical filter is used with 1dB pass-band ripple and 50dB stop-band ripple.

C. Feature Extraction

In this step, five well known features for EEG based research have been considered including temporal features like Hjorth parameters [3], AAR parameters [3], frequency domain features like PSD [3], time-frequency correlated features like discrete wavelet coefficient[3] and CSP features[2]. In this case, AAR parameters of 6-th order have been calculated using Kalman filter as the estimator with an update coefficient of 0.0085. PSD features have been extracted using Welch method with 50% overlapped signal segments using a Hamming window. For wavelet coefficients, Daubechies 4-th order mother wavelet has been used.

D. Optimal Feature and Electrode Selection using Optimization Algorithm

A potential solution of this optimization problem of selection of the optimal set of EEG electrodes [5] and EEG features is encoded as a four dimensional firefly position (Fig. 3). Using a set of such firefly positions, the proposed SAFA is executed. In each generation of SAFA, three steps are followed to determine the objective function value of a firefly position. First, each firefly position is decoded using Fig. 3 to obtain M ($\le N$) and D ($\le F$). The source signals, corresponding to the selected M ($\le N$) electrodes, are estimated using ICA. The selected D ($\le F$) features are then extracted from the selected M ($\le N$) sources and sinks. The fitness of the solution is then determined using (15). Maximization of (15) using FASA ultimately provides the optimal set of electrodes and EEG features for the effective representation of C cognitive tasks i) without dropping out any useful information and ii) without any redundant information.

E. Experimental Results

Table-II reports the mean (standard deviation within parenthesis) of the following measures as obtained by the proposed SAFA-based realization over 50 independent runs, each with 4×10^4 maximum number of function evaluations. The performance metrics include

i) the average of the distance correlation coefficients between the selected D features of the selected M and the original N set of sources,

ii) the average of the mutual information between the selected D features of the selected M and the original N set of sinks, and

ii) the average value of J_B /J_W due to the selected M sources, each with D features

over 50 runs and for all possible combinations of fifteen cognitive tasks. It is evident from Table-II that for emotion-based experiments, frontal and prefrontal electrodes have been chosen as the optimal ones apart from a few other channels. It justifies the effectiveness of the proposed scheme as it is known from elementary neuro-physiological knowledge that frontal and prefrontal brain area are majorly responsible for emotion based task executions.

A comparative analysis of the proposed SAFA with other existing algorithms, including self-adaptive artificial bee colony (SAABC) [12], traditional ABC [9] and traditional differential evolution (DE) [10], are undertaken in Table-III with respect to the mean (standard deviation within parenthesis) objective function values of the optimal solutions (using (15)) over 50 independent runs. All the algorithms commence from the same initial population of size 50. The maximum number of function evaluations for each run of an algorithm is set equal to 50×10^4. Their control parameters are set in a manner to have their individual best performance in the present context after a wide variety of experiments. The parameters of SAFA including $\alpha^{\min}$, β_0, and m are respectively set as 0.2, 1 and 1.5. The *limit* cycle for SAABC and ABC is set to 50. A crossover rate of 0.9 is used in case of DE/current-to-best/1.

The statistical significance level of the difference of the 50 samples of the optimal objective function of any two competitive algorithms is verified by the Wilcoxon rank sum test [11] with a significance level α=0.05. The p-values obtained through the rank sum test between the best algorithm and each of the remaining algorithms over all combinations of cognitive tasks are reported in third brackets in Table-III. Here NA stands for *not applicable* covering the cases of comparing the best algorithm with itself. The null hypothesis of this statistical test is concerned with the equivalent performance of all the competitor algorithms. If the p-value, corresponding to the relative performance analysis of the i-th and j-th algorithms, is less than α, then the respective null hypothesis is rejected. The reported results in Table-III clearly indicate the superiority of the proposed SAFA to its contenders in a statistically significant fashion over most of the combinations of cognitive tasks. SAABC, which yields the best objective function values for three cases (including experiments 7, 11 and 14) attain the second best rank. However, for experiment 11, the statistical test indicates an insignificant dominance of SAABC over SAFA. It is noteworthy that DE based realization of the problem outperforms the proposed SAFA for task 9, however, insignificantly.

It is also remarkable from Tables II and III that the performance of each algorithm remains better for selection of optimal electrodes and EEG features for different classes of motor intensions or motor imagery rather than emotion recognition. An obvious reason may be that the emotional stimuli produce the brain rhythms essentially in α and θ bands, but here the signals are band-pass filtered as a whole in the 4-40 Hz band. It may have resulted in a degraded performance of an algorithm as compared to other cognitive tasks.

The comparative analysis of the competitor algorithms is also undertaken with respect to the classification accuracy of three different SVM classifiers [4], considered in this paper. This is accomplished by first creating a testing dataset for each task combinations in Table-I. The testing dataset is created by following the same principle as in IV.A. Then for an algorithm, say i, we obtain optimal set of M electrodes and D features. Then these D-dimensional M source and sink feature vectors of the training dataset are used to train the SVM individually. After completing the training cycle, the D-dimensional M source and sink feature vectors are extracted from the testing set and are used to verify the classification accuracy of the trained SVM. This is repeated for all four competitors with i=[1, 4] for all three variants of nonlinear SVM. Fig. 6 plots the classification accuracy obtained by individual SVM classifier due to selection of optimal set of EEG electrodes and features by each contender algorithm. The plot clearly reveals that the proposed SAFA here too outperforms its contenders, however, marginally for SAABC [12].

TABLE II-A: PERFORMANCE OF SAFA BASED SELECTION OF OPTIMAL EEG ELECTRODES AND FEATURES FOR COGNITIVE TASK CLASSIFICATION FOR EXPERIMENTS 1 TO 7

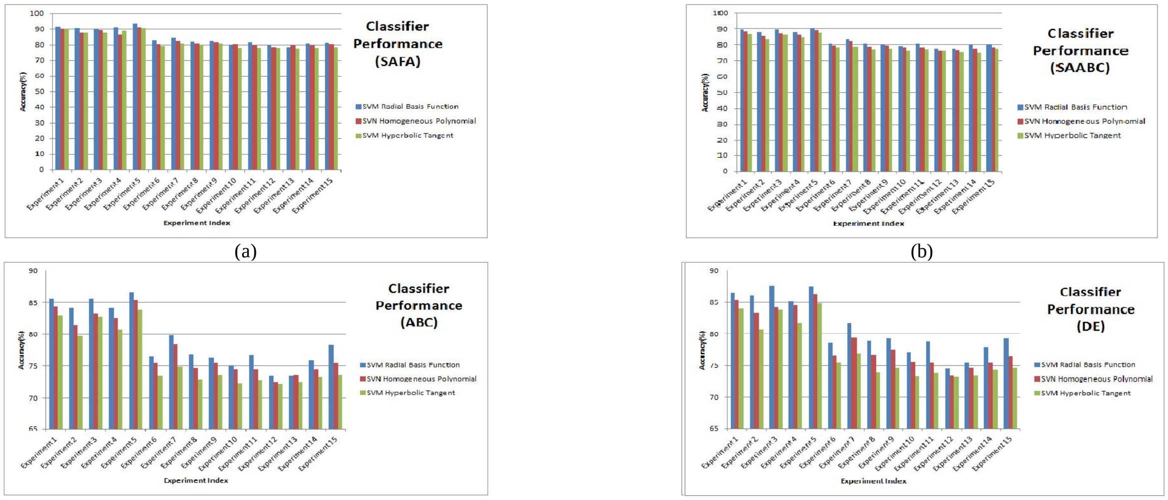
Experiment Index	Number of Selected Electrodes	Optimal Electrode Positions	Number of Selected Features	Correlation	J_B/J_W	Mutual Information
1	5	C3, C4, Cz, P4, P3	545	320.51 (0.047)	11.09 (0.068)	12.66 (0.043)
2	4	C3, C4, Cz, Pz	412	288.75 (0.013)	9.43 (0.027)	10.06 (0.031)
3	6	C3, C4, Cz, P3, P4, Pz	755	373.89 (0.057)	18.23 (0.024)	20.65 (0.075)
4	3	P3, P4, Pz	323	180.76 (0.062)	8.39 (0.061)	8.85 (0.028)
5	5	C3, C4, P3, P4, Pz	563	331.87 (0.023)	16.51 (0.053)	16.78 (0.011)
6	4	FP1, FP2, Cz, F4	472	270.57 (0.039)	12.64 (0.038)	11.57 (0.027)
7	4	FP1, C3, Fz, F3	397	225.64 (0.032)	8.57 (0.014)	9.68 (0.017)

TABLE II-B: PERFORMANCE OF SAFA BASED SELECTION OF OPTIMAL EEG ELECTRODES AND FEATURES FOR COGNITIVE TASK CLASSIFICATION FOR EXPERIMENTS 8 TO 15

Experiment Index	Number of Selected Electrodes	Optimal Electrode Positions	Number of Selected Features	Correlation	JB /JW	Mutual Information
8	5	FP1, FP2, Fz, F3, C3	513	321.65 (0.076)	11.67 (0.032)	10.89 (0.072)
9	6	FP1, F3, F4, O1, C3, Cz	781	365.45 (0.083)	17.85 (0.011)	18.95 (0.069)
10	4	T3, FP1, F3, F4	456	265.88 (0.028)	8.58 (0.045)	9.98 (0.083)
11	5	O_2, FP2, Fz, F3, F4	529	318.77 (0.048)	11.96 (0.087)	13.46 (0.071)
12	3	FP1, FP2, Fz	293	200.45 (0.022)	7.43 (0.069)	8.93 (0.059)
13	5	T3, FP1, F3, F4, FP2	527	322.68 (0.059)	10.08 (0.052)	12.89 (0.081)
14	6	FP1, Pz, Fz, T3, F4, T4	746	354.05 (0.017)	18.79 (0.047)	19.58 (0.079)
15	4	T3, FP1, FP2, F4	468	278.92 (0.086)	11.23 (0.033)	11.87 (0.066)

TABLE III: COMPARISON OF THE PROPOSED ALGORITHM WITH OTHER STANDARD EVOLUTIONARY ALGORITHMS BASED ON OBJECTIVE FUNCTION VALUES

Experiment Index	SAFA	SAABC	ABC	DE
1	**0.2562 (0.056) [NA]**	0.2068 (0.049) [0.0034]	0.0167 (0.011) [0.0053]	0.1583 (0.027) [0.0016]
2	**0.2963 (0.095) [NA]**	0.2477 (0.064) [0.0048]	0.0543 (0.031) [0.0074]	0.1948 (0.072) [0.0031]
3	**0.3549 (0.025) [NA]**	0.3249 (0.112) [0.0073]	0.0851 (0.105) [0.0084]	0.2057 (0.087) [0.0027]
4	**0.3688 (0.145) [NA]**	0.2911 (0.194) [0.0046]	0.1049 (0.162) [0.0033]	0.2648 (0.121) [0.0051]
5	**0.3344 (0.271) [NA]**	0.3105 (0.127) [0.0053]	0.1078 (0.186) [0.0082]	0.2420 (0.191) [0.0154]
6	**0.3231 (0.084) [NA]**	0.2865 (0.073) [0.0049]	0.0989 (0.085) [0.0064]	0.1997 (0.066) [0.0129]
7	0.3825 (0.094) [0.0044]	**0.3941 (0.105) [NA]**	0.2732 (0.147) [0.0018]	0.3475 (0.194) [0.0003]
8	**0.3376 (0.034) [NA]**	0.2951 (0.072) [0.0059]	0.0611 (0.128) [0.0011]	0.2559 (0.136) [0.0046]
9	0.2931 (0.241) [0.0564]	0.2883 (0.262) [0.0077]	0.2691 (0.177) [0.0003]	**0.3259 (0.167) [NA]**
10	**0.3596 (0.184) [NA]**	0.3115 (0.179) [0.0165]	0.1572 (0.086) [0.0234]	0.2754 (0.109) [0.0196]
11	0.3104 (0.097) [0.0602]	**0.3155 (0.116) [0.0007]**	0.2158 (0.123) [0.0097]	0.2795 (0.152) [0.0197]
12	**0.3963 (0.096) [NA]**	0.3386 (0.188) [0.0217]	0.1652 (0.192) [0.0194]	0.2964 (0.271) [0.089]
13	**0.3713 (0.290) [NA]**	0.3494 (0.306) [0.0204]	0.1922 (0.327) [0.0028]	0.2278 (0.315) [0.0011]
14	0.3353 (0.224) [0.0189]	**0.3686 (0.219) [0.0173]**	0.2263 (0.254) [0.0184]	0.3281 (0.278) [0.008]
15	**0.3533 (0.172) [NA]**	0.3467 (0.245) [0.0309]	0.0476 (0.239) [0.0061]	0.2343 (0.241) [0.0149]



(a) (b) (c) (d)

Fig. 6 Classification accuracy of SVM classifiers dueto training with EEG electrodes and features selected by (a) SAFA), (b) SAABC), (c) ABC and (d) DE

V. Conclusion

The paper proposes a novel evolutionary optimization approach to the simultaneous selection of significant electrodes and relevant EEG features for classification of cognitive tasks in EEG based BCI paradigm. The present work can get extreme appreciation, because till date separate techniques of electrode and feature selections are adopted by the researchers, but a combined methodology of simultaneous dealing with these two important aspects of EEG signal processing has not been used in large scale. It is noteworthy that the efficiency of classification of cognitive tasks is enhanced by a great extent using the proposed framework. The main novelty of the work lies in the formulation of the problem in an optimization framework and in solving the problem using the proposed SAFA with an aim to satisfy three criteria. 1) To minimize the loss of information of the cortical sources for a cognitive task by discarding a specific set of redundant electrodes, the correlation between the EEG sources before and after the electrode selection should be as high as possible (in the feature space). 2) To ensure that only relevant EEG electrodes to be used for characterizing the mental states of cognitive tasks, the mutual information between the selected electrodes should be as low as possible indicating their independence (in the feature space). 3) The optimal set of features is selected by identifying the features that well differentiate between EEG sources of two cognitive tasks, while representing a specific cognitive task uniquely. To analyze the relative performance of SAFA with other existing evolutionary algorithms, three variants of non-linear SVM [4] classifier (including Gaussian radial basis function, homogeneous polynomial function and hyperbolic tangent function) are individually trained with the selected feature sets of the selected EEG electrodes and the corresponding EEG sources, as obtained by individual competitor algorithms. Experiments undertaken with SAFA reveal the statistically significant superiority of the proposed method over other existing evolutionary algorithms, with respect to classification accuracy of the SVM classifier, irrespective of the non-linearity.

References

[1] Blankertz, B.; Tomioka, R.; Lemm, S.; Kawanabe, M.; Muller, K.-R., "Optimizing Spatial filters for Robust EEG Single-Trial Analysis," in Signal Processing Magazine, IEEE , vol.25, no.1, pp.41-56, 2008.

[2] Yijun Wang; Shangkai Gao; Xiaorong Gao, "Common Spatial Pattern Method for Channel Selelction in Motor Imagery Based Brain-computer Interface," in Engineering in Medicine and Biology Society, 2005. IEEE-EMBS 2005. 27th Annual International Conference of the , vol., no., pp.5392-5395, 17-18 Jan. 2006.

[3] Datta, S.; Rakshit, P.; Konar, A.; Nagar, A.K., "Selecting the optimal EEG electrode positions for a cognitive task using an Artificial Bee Colony with Adaptive Scale Factor optimization algorithm," in Evolutionary Computation (CEC), 2014 IEEE Congress on , vol., no., pp.2748-2755, 6-11 July 2014.

[4] Saha, A.; Konar, A.; Das, P.; Bhattacharya, B.S.; Nagar, A.K., "Data-point and feature selection of motor imagery EEG signals for neural classification of cognitive tasks in car-driving," in Neural Networks (IJCNN), 2015 International Joint Conference on , vol., no., pp.1-8, 12-17 July 2015.

[5] Saha, A.; Konar, A.; Rakshit, P.; Ralescu, A.L.; Nagar, A.K., "Olfaction recognition by EEG analysis using differential evolution induced Hopfield neural net," in Neural Networks (IJCNN), The 2013 International Joint Conference on , vol., no., pp.1-8, 4-9 Aug. 2013.

[6] Yang, Xin-She. "Firefly algorithms for multimodal optimization." In *Stochastic algorithms: foundations and applications*, pp. 169-178. Springer Berlin Heidelberg, 2009.

[7] Székely, Gábor J., Maria L. Rizzo, and Nail K. Bakirov. "Measuring and testing dependence by correlation of distances." *The Annals of Statistics* 35, no. 6 (2007): 2769-2794.

[8] Lin, Tung-Po. "The power mean and the logarithmic mean." *The American Mathematical Monthly* 81, no. 8 (1974): 879-883.

[9] Basturk, Bahriye, and Dervis Karaboga. "An artificial bee colony (ABC) algorithm for numeric function optimization." In *IEEE swarm intelligence symposium*, vol. 8, no. 1, pp. 687-697. 2006.

[10] Storn, Rainer, and Kenneth Price. "Differential evolution–a simple and efficient heuristic for global optimization over continuous spaces." *Journal of global optimization* 11, no. 4 (1997): 341-359.

[11] Sheskin, David J. *Handbook of parametric and nonparametric statistical procedures*. crc Press, 2003.

[12] W. Gu, M. Yin, and C. Wang, “Self Adaptive Artificial Bee Colony for Global Numerical Optimization,” IERI Procedia 1, pp. 59-65, 2012.